

a display screen, including touch display screens or multi-touch display screens, keyboard or other input devices, microphones, speakers, etc. In other embodiments, the system 150 may be any type of computing system (e.g. desktop personal computer, laptop, workstation, net top etc.).

[0055] The external memory 158 may include any type of memory. For example, the external memory 158 may be SRAM, dynamic RAM (DRAM) such as synchronous DRAM (SDRAM), double data rate (DDR, DDR2, DDR3, etc.) SDRAM, RAMBUS DRAM, low power versions of the DDR DRAM (e.g. LPDDR, mDDR, etc.), etc. The external memory 158 may include one or more memory modules to which the memory devices are mounted, such as single inline memory modules (SIMMs), dual inline memory modules (DIMMs), etc. Alternatively, the external memory 158 may include one or more memory devices that are mounted on the IC 152 in a chip-on-chip or package-on-package implementation.

[0056] FIG. 7 is a block diagram of one embodiment of a computer accessible storage medium 160 storing an electronic description of the IC 152 (reference numeral 162) is shown. More particularly, the description may include at least the matrix computation engine 10 and/or the processor 12. Generally speaking, a computer accessible storage medium may include any storage media accessible by a computer during use to provide instructions and/or data to the computer. For example, a computer accessible storage medium may include storage media such as magnetic or optical media, e.g., disk (fixed or removable), tape, CD-ROM, DVD-ROM, CD-R, CD-RW, DVD-R, DVD-RW, or Blu-Ray. Storage media may further include volatile or non-volatile memory media such as RAM (e.g. synchronous dynamic RAM (SDRAM), Rambus DRAM (RDRAM), static RAM (SRAM), etc.), ROM, or Flash memory. The storage media may be physically included within the computer to which the storage media provides instructions/data. Alternatively, the storage media may be connected to the computer. For example, the storage media may be connected to the computer over a network or wireless link, such as network attached storage. The storage media may be connected through a peripheral interface such as the Universal Serial Bus (USB). Generally, the computer accessible storage medium 160 may store data in a non-transitory manner, where non-transitory in this context may refer to not transmitting the instructions/data on a signal. For example, non-transitory storage may be volatile (and may lose the stored instructions/data in response to a power down) or non-volatile.

[0057] Generally, the electronic description 162 of the IC 152 stored on the computer accessible storage medium 160 may be a database which can be read by a program and used, directly or indirectly, to fabricate the hardware comprising the IC 152. For example, the description may be a behavioral-level description or register-transfer level (RTL) description of the hardware functionality in a high level design language (HDL) such as Verilog or VHDL. The description may be read by a synthesis tool which may synthesize the description to produce a netlist comprising a list of gates from a synthesis library. The netlist comprises a set of gates which also represent the functionality of the hardware comprising the IC 152. The netlist may then be placed and routed to produce a data set describing geometric shapes to be applied to masks. The masks may then be used in various semiconductor fabrication steps to produce a

semiconductor circuit or circuits corresponding to the IC 152. Alternatively, the description 162 on the computer accessible storage medium 300 may be the netlist (with or without the synthesis library) or the data set, as desired.

[0058] While the computer accessible storage medium 160 stores a description 162 of the IC 152, other embodiments may store a description 162 of any portion of the IC 152, as desired (e.g. the matrix computation engine 10 and/or the processor 12, as mentioned above).

[0059] Numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

What is claimed is:

1. An apparatus comprising:

a processor configured to fetch a first instruction; and  
a computation engine coupled to the processor, wherein:

the computation engine comprises at least two input memories configured to store input vectors of elements and an output memory configured to accumulate an output vector of elements; and

the computation engine comprises a multiply accumulation (MAC) circuit coupled to the at least two input memories and configured to perform a multiply-accumulate on the input vectors of elements responsive to the first instruction, generating the output vector of elements for the output memory, wherein the MAC circuit is configured to perform multiplications on elements having a first size;

the computation engine comprises a lookup table programmed to map first elements of a second size less than the first size to second elements of the first size, wherein the first elements are provided from at least one of the input memories, and wherein the second elements are provided to the MAC circuit.

2. The apparatus as recited in claim 1 wherein the lookup table is programmable from the processor.

3. The apparatus as recited in claim 2 the lookup table is stored in one of the at least two input memories.

4. The apparatus as recited in claim 1 wherein the MAC circuit comprises a plurality of MACs, each MAC configured to generate one element in the output vector.

5. The apparatus as recited in claim 1 wherein the input vectors and output vectors are vectors of matrices, and wherein the MAC circuit comprises a plurality of MACs, each MAC configured to generate one matrix element of one matrix in the output vector.

6. The apparatus as recited in claim 5 wherein a number of the plurality of MACs is equal to a number of matrix elements in the output vector.

7. The apparatus as recited in claim 5 wherein a number of the plurality of MACs is less than a number of matrix elements in the output vector, and wherein the first instruction is completed by iterating the plurality of MACs with different portions of the input vectors and the output vector.

8. A matrix computation engine comprising:

a circuit configured to perform a matrix multiplication operation on a first vector operand and a second vector operand, producing a resulting output vector;

wherein the circuit is configured to operate on matrix elements having a second size greater than a first size; and